Improving Brain Tumor MRI Image Classification Prediction based on Fine-tuned MobileNet

Quy Thanh Lu, Triet Minh Nguyen, Huan Le Lam
Information Technology Department, FPT University, Can Tho, Viet Nam

***Abstract*—Brain tumors are a prevalent issue in contemporary society as they impact human health. The location of the tumor in the brain determines the variety of symptoms that may manifest. Some frequent symptoms are cephalalgia, convulsions, visual impairments, nausea, emesis, asthenia, paresthesia, dysphasia, personality alterations, and amnesia. The prognosis for brain cancer differs considerably depending on the cancer type. Nevertheless, brain tumors are amenable to treatment with surgical intervention, chemotherapy, and radiotherapy if the diagnosis is timely. Furthermore, artificial intelligence and machine learning can assist in the detection of brain tumors as they have significant implications for the analysis of Magnetic Resonance Imaging (MRI). To accomplish this objective, automated measurement instruments were proposed based on the processing of MRI. In this study, we employed the latest developments in deep transfer learning and fine-tuning to identify tumors without many complex steps. We gathered data from authentic MRI of 3264 subjects (i.e., 926 glioma tumors, 937 meningioma tumors, 901 pituitary tumors, and 500 normal). With the MobileNet model from the Keras library, we attained the highest validation accuracy, test accuracy, and F1 score in four-class classifications was 97.24%, 97,86%, and 97.85%, respectively. Concerning two-class classification, high accuracy values were obtained for most of the models (i.e., ~100%). These outcomes and other performance indicators demonstrate a strong capability to diagnose brain tumors from conventional MRI. The current research developed a supportive machine learning that can aid doctors in making the accurate diagnosis with less time and mistakes.**

Keywords*—Brain tumor; fine-tuning; transfer learning; Magnetic Resonance Imaging (MRI); MobileNet

I. Introduction

The health of people in the modern world is adversely affected by numerous diseases, among which brain tumors are prevalent in various age groups, from adolescents to seniors. Brain tumors are not a helpless disease and there are several treatment options available for patients with this disease. Some of the most common and efficacious methods of treating brain tumors are surgery, chemotherapy, radiation therapy, or a combination of these techniques. Surgery entails excising the tumor or part of it through an operation, while chemotherapy and radiation therapy employ drugs and high-energy rays to eradicate or reduce the tumor cells. The selection of treatment for each patient hinges on multiple factors, such as the type of the tumor, the grade of the tumor, the location of the tumor, the size of the tumor, as well as the age and general health of the patient. By considering these factors, physicians can devise the optimal treatment plan for each patient and enhance their prospects of recovery.

Nevertheless, it is still dangerous if we avoid it. For example, A brain tumor is a grave condition irrespective of its benignity or malignancy. Tumors within the cranium expand and exert pressure on different regions of the brain, impairing their function. A glioma is a tumor that emerges when glial cells proliferate abnormally. Typically, these cells support neurons and facilitate the operation of your central nervous system. Gliomas commonly grow in the brain, but can also emerge in the spinal cord. Gliomas are malignant (cancerous), but some can be very slow-growing. In addition, Meningioma is the most frequent type of primary brain tumor, it is a tumor that develops from the meninges, the protective layers around the brain and spinal cord. Meningiomas occur more often in women and are usually detected at older ages. Moreover, Pituitary tumors are abnormal enlargements that originate in the pituitary gland. Some of these tumors cause the pituitary gland to produce an excess of certain hormones that regulate vital body functions. tumors stimulate the adrenal glands to produce too much cortisol. This causes a condition called Cushing disease. Others can cause the pituitary gland to produce too little of those hormones.

Numerous data gathered in recent years globally indicate the detrimental effects of brain tumors. The CBTRUS Statistical Report states that from 2016 to 2020, 86,030 deaths occurred due to malignant brains. This corresponds to an average annual mortality rate of 4.42 and an average of 17,206 deaths per year [1]. In 2030, 145,650 new cases of brain tumors are projected to be diagnosed in China with 68,730 men and 76,920 women [2]. The 5-year relative survival rate after diagnosis of a malignant brain tumor was 35.7% and for a non-malignant brain tumor was 91.9% [1]. The most frequent malignant brain tumor was glioblastoma with 14.2% of all tumors, and the most common non-malignant tumor was meningioma with 40.8% of all tumors [1]. Glioblastoma was more prevalent in males, and meningioma was more prevalent in females. In children and adolescents aged 0-19 years, the incidence rate of primary tumors was 6.14 out of 100,000 people [3]. An estimated 3,920 new cases of primary childhood brain tumors are expected to be diagnosed in 2023 [4]

Recently, Magnetic Resonance Imaging (MRI) has been considered one of the most efficient methods for identifying the irregular parts of the central nervous system and human brain. Moreover, the early diagnosis of brain tumors is one of the crucial tasks and offers many advantages to patients. Early identification of tumors helps doctors devise suitable treatment plans and helps lower mortality in patients with brain tumors as quickly as possible. There are many methods that clinicians can use to detect brain tumors as Computerized Tomography

(CT) or MRI. However, it would be time-intensive to diagnose using the MRI method. Furthermore, doctors have to prepare and conduct many procedures to finish a standardized process for patients. Therefore, applying the newest advanced techniques in the diagnostic process can spare valuable time for doctors. Moreover, it can provide doctors with a recommendation to improve the diagnostic process and increase the outcomes.

Artificial intelligence is one of the most prominent advanced techniques invented in recent years. Therefore, we decided to use it to assist doctors in diagnosing MRI images. In the field of artificial intelligence, machine learning is related to the development and research of statistical algorithms that can effectively generalize data and syntax. Based on the trained and prepared results, the computer will perform tasks without explicit instructions [5]. Machine learning algorithms have many applications. For example, financial prediction, transportation, education, data structure in health care systems, drug reaction prediction, diabetes research, cyber security, banking and finance, and social media [6].

Our study used transfer learning and fine-tuning, a part of deep learning which is the subset of the machine learning method. That aim enables us to reuse pre-trained models for new tasks and datasets. Moreover, Transfer learning concentrates on transferring general knowledge from one domain to freeze certain layers to preserve general [7]. Fine-tuning adapts the model to a particular allows the pre-trained layers to be updated [8]. We suggest a new method to use the MobileNet model in the Keras library in a Convolutional Neural Network (CNN) which is appropriate for image recognition and processing tasks and it has achieved state-of-the-art outcomes on a wide range of image recognition tasks, such as object classification, object detection, and image segmentation [9]. It is trained using a huge dataset of annotated images. Once trained, a CNN can be utilized to classify new images or extract features for use in other applications such as object detection or image segmentation.

The contributions of this paper are as follows:

- We propose a complete reliable artificial intelligence model that is used for brain tumor classification including glioma, meningioma, and pituitary tumor. Hence, it can create an easy and speedy way for doctors to detect and classification on their medical purpose.
- Our method achieves a high performance (i.e., <97%) of the deep learning models for four-class (glioma, meningioma, pituitary, and normal) and pair-wise classification problems also achieve a high success (i.e., ~100%). As a result. The effectiveness of each model in terms of training and testing times was also evaluated.
- Our collected MRIs of subjects afflicted with brain tumors, as well as healthy ones, as verified by the specialists in the hospital. This dataset is confirmed for the development of automated machine learning and AI algorithms for the detection of brain tumors and can be applied to educating medical students.
- We point out that GradCam can be used effectively for visual explanations. So, highlighting the important regions (i.e., glioma tumor, meningioma tumor, and pituitary tumor) in the image can provide a more intuitive view for physicians

Our research paper comprises four main sections. In the subsequent Section II, we indicate some of the related research that we employed for references. Following the related research section is the methodology Section III, this section elucidates in detail all of the methods utilized in the article. Subsequently, Section IV will refer to the experiments, and how we conduct and evaluate the accuracy of the deep learning model. Lastly, in the final Section V, we summarize our article and examine the essential domains associated with the study.

II. Related Work

Besides the change in environment and population, a lot of diseases have negative effects on human life and one of these is brain tumors. The survey pointed out that 1,323,121 people living with brain and other CNS tumors (malignant and non-malignant) on December 31, 2019 [1]. At that time, a bunch of research on medical and artificial intelligence helped humans in the healing of those sicknesses. Wadhah Ayadi detected brain tumors by suggesting a new CNN that contains various layers such as convolution, Rectified Linear Unit (Relu), and pooling to achieve a best Accuracy is 94.74 [10]. Following Ahmad Saleh, His scholarly investigation endeavors to enhance the proficiency of MRI apparatus in the categorization of cerebral neoplasms and discerning their respective classifications, using AI Algorithms, CNN, and Deep Learning. The MobileNet model is used in this study. The evaluation of the image dataset is conducted utilizing the F1-score metric, yielding a commendable accuracy rate of 97.25% [11].

Machine Learning helped experts and doctors research medical image analysis. This led to, a change in normal examination and treatment to aid medical procedures to avoid a waste of time and money. Hence, typical studies appear more and more such as Wadhah Ayadi. The presented methodology used normalization, dense speeded-up robust features, and histogram of gradient techniques to enhance the quality of MRI and produce a discriminative feature set. The accuracy attained through the implementation of this method is measured at 90.27% [12]. In addition, Muhammad Imran Sharif suggests Densenet201 Pre-Trained Deep Learning Model. The attributes of the trained model are derived from the average pooling layer, elucidating the profound information on each specific tumor type. In addition, He includes two new models Entropy–Kurtosis-based High Feature Values (EKbHFV) and modified genetic algorithm (MGA). Finally, the research paper has achieved an accuracy higher than 95% [13].

Convolutional Neural Network (CNN) is one of the parts of deep learning models commonly used in Computer Vision that help computers understand and interpret images or visual data. It has main contributions to creating intelligent systems with great accuracy. So, it has produced several academic works as S Kumar's research paper. In short, the technical use of the Deep Convolution Neural Network (Deep CNN) for performing the brain tumor classification with Dolphin-SCA as the training algorithm. The database is MRI images given by the BRATS database and SimBRATS, and the suggested model has shown a maximum accuracy of 96.39% [14]. Moreover, Díaz-Pernas applied the method on a publicly available MRI image dataset of 3064 images from 233 patients compared with previously classical segmentation and classification published methods. In this comparative analysis, the proposed method demonstrated exceptional results, achieving an impressively high tumor classification accuracy of 0.973 [15].

Processing medical images by automatic segmentation and classification becoming extremely important around the world [16], [17], especially in the medical field such as diagnostics, growth prediction, and treatment of brain tumors. As a result, a patient can save their life because an early detection of brain tumors that helps to increase their survival rate. Applying machine learning, the brain classification paper from Huong Hoang Luong points out that the K-mean clustering algorithm stratifies the samples into three distinct view angles of MRI, namely, transverse, coronal, and sagittal planes. This strategic classification process was coupled with the integration of a modified Residual Network (ResNet) architecture. Finally, He reached a performance of 96% in the brain tumor classification accuracy [18]. Furthermore, with another k-means algorithm S. Rinesh's innovative approach outperformed conventional methods such as hybrid k-means clustering and parallel k-means clustering, exhibiting superior results with a higher peak signal-to-noise ratio and a reduced mean absolute error value. The proposed model attained an accuracy of 96.47% [19].

Additionally, we have introduced an automated classification system designed for the intricate challenge of categorizing multiclass brain tumor MRI. This task, inherently more complex and demanding, transcends the relative simplicity of binary classification. The dataset is almost equal to Khan Swatithe's paper uses a pre-trained deep CNN model and introduces a block-wise fine-tuning strategy rooted in transfer learning principles. Notably, this methodology is characterized by its generality, eschewing the need for handcrafted features, and demanding minimal preprocessing. Impressively, the proposed approach attains an average accuracy rate of 94.82% within the context of a five-fold cross-validation framework [20]. Besides, with a straightforward architectural design and the absence of any antecedent region-based segmentation, Nyoman Abiwinanda achieved commendable results, attaining a training accuracy of 98.51% and a peak validation accuracy of 84.19%. Notably, these outcomes stand in favorable comparison to the performance of more intricate region-based segmentation algorithms [21].

In summary, the related work saw notable weaknesses, particularly characterized by low accuracy. Many current models struggle to consistently achieve high precision across diverse datasets, leading to concerns about their robustness and generalizability. Additionally, the lack of a visual explanation such as GradCam for evaluating and comparing models hinders progress. Addressing these limitations is imperative to propel deep learning toward more reliable and universally applicable solutions, ensuring advancements that transcend the current constraints of accuracy and evaluation methodologies.

III. METHODOLOGY

A. The Research Implementation Procedure

In this research paper, we propose a method including 11 steps from input to output shown in Fig. 1. The roles of the steps are shown as follows:

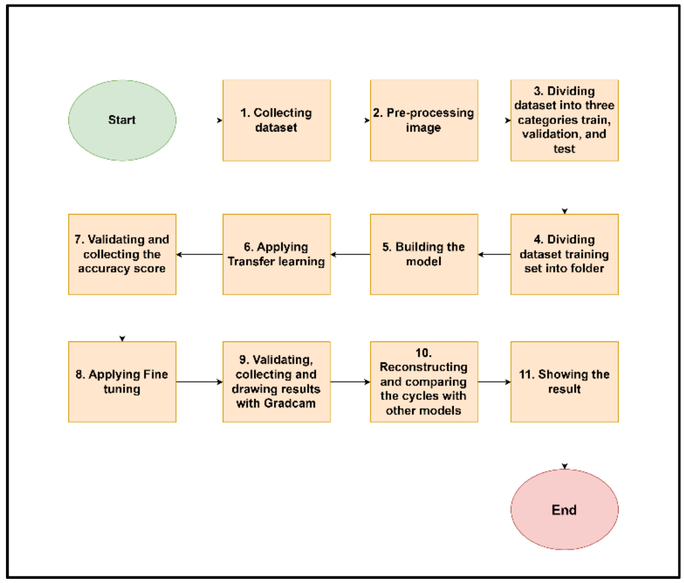

Fig. 1. The implementing procedure flowchart.

1) Collecting dataset: The dataset is collected by Swati Kanchan in B. Tech Dept of CSE from NIT Durgapur. The dataset comprises MRI images, including three types of brain tumors—meningioma, glioma, and pituitary—as well as normal images. This comprehensive collection serves as a valuable resource for medical research and diagnostic advancements.

2) Pre-processing image: This important step requires both resizing and normalization to establish standardized input conditions for machine-learning models. This leads to an increase in the outcome of the results.

3) Dividing the dataset into three categories train validation and test: In total, 3264 images constitute the comprehensive MRI dataset, with random selection for training, validation, and testing phases. The datasets are randomly chosen using an 8-1-1 scale, allocating 8 parts for training, 1 for validation, and 1 for testing, ensuring a balanced distribution for robust model development and evaluation.

4) Dividing the training set into folders: Types of brain tumors (i.e., meningioma, glioma, and pituitary and normal) are divided into many different folders. The biggest include 4 classes (i.e., meningioma, glioma, and pituitary and normal). We want to show the training data more exactly way and use it to compare the biggest folder. Dense, the other folder contains two class classifications includes: meningioma-normal, glioma-normal, and pituitary-normal

5) Building the model: To conduct experiments, we reconstructed the model based on the prototype of the CNN architecture by inheriting its core processing layers and modifying some layers to achieve improved results. As a result, the model creates an excellent result for our training test with Keras's models library.

6) Applying transfer learning: In transfer learning, the model gets trained on a big set of data for a specific job. This dataset could be general or have lots of labeled information. The things the model learns are saved in its weights, especially in the lower layers. These layers gather important features from the input data, making it easier for the model to understand and work with the information.

7) Validating and collecting the accuracy score: After completing the training of the model, we assessed its precision by summarizing the training accuracy derived from the model's predictions. Subsequently, we evaluated the accuracy of the test set using the initially separated testing set."

8) Applying fine tuning: In this procedure, the process of adjusting the hyperparameters of a model to improve its performance on a dataset and using the weights of an already trained model as the starting values for training a new model. Hyperparameters are the parameters that control the learning process and we finish after the model has been trained on a training set, and done when using a validation set.

9) Validating, collecting and drawing results with GradCam: We summary all the metrics like validation accuracy, test accuracy, and F1 score, then. Images used GradCam to represent detected brain tumors in color to more accurately represent the results.

10)Reconstructing and comparing the cycles with other models: When we have results in one model, we rework another model in Keras including MobileNet, Inception V3, VGG16, ResNet50, and EfficientNet B3 to compare the final result

11)Showing the result: After conducting a comparison, results will be presented through tables and graphs to facilitate meaningful comparisons.

B. Pre-processing Image

In the pre-processing pipeline for images, a pivotal step involves both resizing and normalization to establish standardized input conditions for machine learning models. The resizing operation transforms the input image I to a uniform dimension of (new width, new height). As a result, we decided to choose 224 pixels for weight and 224 pixels for height (1) using a chosen interpolation method, as captured by the formula:

$$I_{resize}\left(new_{width}, new_{height}\right) = Resize(I, 244,244) \quad (1)$$

This operation (1) is essential for establishing a consistent input size, a prerequisite in various deep-learning applications. Following resizing in equation (2), the subsequent normalization process adjusts pixel values to a range between 0 and 1, facilitating model training and convergence. The normalization is mathematically expressed as:

$$O(x, y) = \frac{I_{resize}(x,y)}{255} \quad (2)$$

In this Eq. 2, $O(x, y)$ signifies the output pixel value at position (x, y) in the preprocessed image. The division by 255 ensures that the pixel values are scaled to fit within the [0, 1] range, aligning with common conventions in image processing.

This dual procedure of resizing to (224, 224) (1) and subsequent normalization not only ensures a standardized size for all images but also provides a consistent pixel value scale (2), thereby enhancing the efficiency and robustness of downstream machine learning tasks.

C. Transfer Learning and Fine Tuning of CNN (MobileNet)

We noticed that transfer learning and fine-tuning of CNN have become pivotal techniques in the realm of computer vision, enabling the effective utilization of pre-trained models on new tasks [22][23][24]. One such exemplary model is MobileNet, a lightweight and efficient architecture design. MobileNet, characterized by depth wise separable convolutions, significantly reduces the computational burden while preserving the model's capacity to capture intricate features in images.

The MobileNet model's architecture is inherently rooted in the principles of convolutional neural networks, with its convolutional layers serving as feature extractors. This architecture excels in tasks demanding real-time performance and resource efficiency. When considering transfer learning, MobileNet offers a valuable starting point. Pre-trained on large-scale image datasets, it possesses a robust understanding of general image features, making it an ideal candidate for various computer vision applications.

The integration of MobileNet into the broader CNN architecture is seamless and complementary such as Fig. 2. The initial layers of a CNN, responsible for low-level feature extraction, can be substituted with the MobileNet backbone. This strategic replacement allows the model to retain its ability to recognize high-level features while benefiting from MobileNet's efficiency in processing low-level features. The resulting hybrid architecture strikes a balance between computational efficiency and task-specific adaptability.

In conclusion, the fusion of transfer learning and fine-tuning within the realm of CNNs, particularly with the utilization of the MobileNet model, represents a powerful approach to solving diverse computer vision challenges. This leads to the final results about validation accuracy, test accuracy, and F1 score increase and reaching a desirable point.

D. Visual Explanation by GradCam

In our paper, we decided to choose a visual explanation by Gradient-weighted Class Activation Mapping (i.e., GradCam), because it stands as a pivotal technique in unraveling the decision-making processes of CNN. It operates by shedding light on crucial regions within an image that heavily influence the model's final prediction, thus enhancing both interpretability and confidence in the outputs.

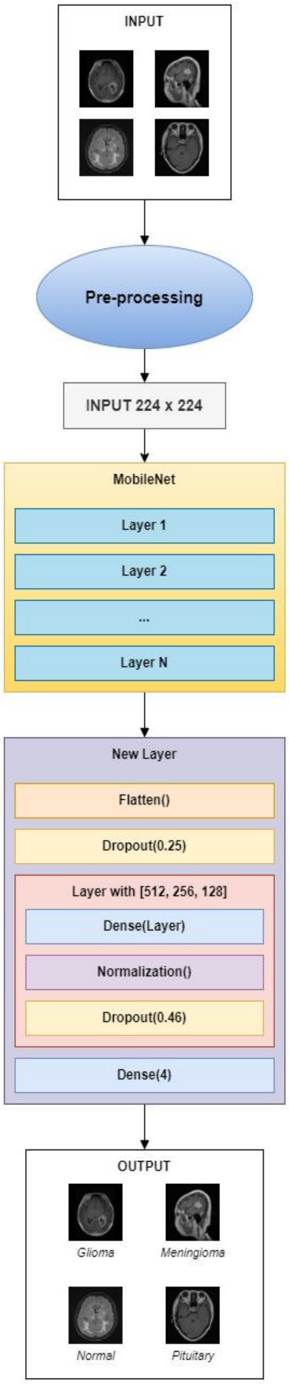

Fig. 2. Procedure of transfer learning and fine-tuning in ours model with custom layers.

In more detail, GradCam relies on the gradient information flowing into the final convolutional layer of a CNN. Let F^k (3) represent the k-th feature map from the last convolutional layer, and w_c^k (3) show the weight of the k-th feature map for the target class 'c.' The GradCam heatmap $L_c^{GradCam}$ (3) is computed as the global average pooling of the positive gradients:

$$L_c^{GradCam} = ReLU(\sum_k w_c^k \cdot F^k) \tag{3}$$

The ReLU (3) function ensures that only positive contributions are considered, highlighting the regions with a positive influence on the decision for the target class 'c' Fig. 3.

To explain this formula in detail, let's express the weight w_c^k (4) for a specific feature map 'k' as the global average pooling of the gradients:

$$w_c^k = \frac{1}{Z} \sum_{i,j} \frac{\partial y_c}{\partial A_{i,j}^k} \tag{4}$$

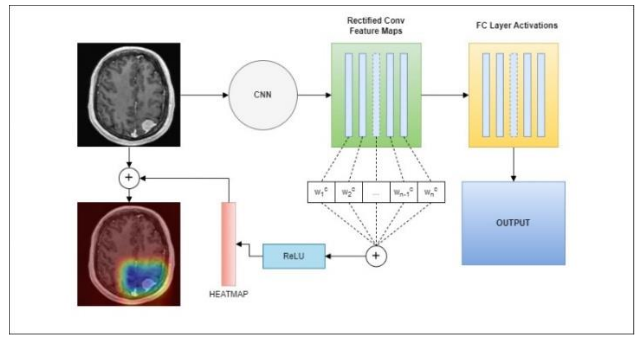

Fig. 3. GradCam functionality.

Because y_c (4) denotes the logit for the target class 'c,' and A_j^k (4) presents the activation of the k-th feature map at position (i, j). The normalization term 'Z' ensures that the weights sum to 1, providing a meaningful contribution measure.

The computation of the gradient $\frac{\partial y_c}{\partial A_j^k}$ (5) involves back-propagating the derivative of the logit with respect to the activation of the k-th feature map at position (i, j). By means of mathematics, this can be expressed as:

$$\frac{\partial y_c}{\partial A_j^k} = \frac{\partial y_c}{\partial F_{i,j}^k} \cdot \frac{\partial F_{i,j}^k}{\partial A_{i,j}^k} \tag{5}$$

This chain of derivatives involves the gradient of the logit with respect to the activation of the k-th feature map at position (i, j), $\frac{\partial y_c}{\partial F_{i,j}^k}$ (5), and the gradient of the activation at (i, j) with respect to the input activation, $\frac{\partial F_{i,j}^k}{\partial A_{i,j}^k}$ (5).

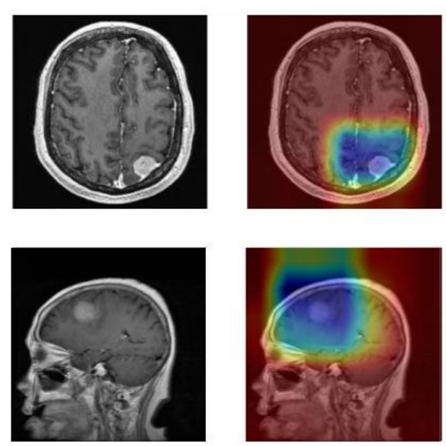

Fig. 4. The final result after applying a visual explanation by GradCam.

By using this GrandCam, our results have the contribution of each feature map to the final decision in Fig. 4, and facilitate future use by professionals and doctors.

IV. Experiments

A. Dataset and Peformance Metrics

The research used a single dataset for both the training, validation, and testing phases in this analysis. The data was taken and augmented by Swati Kanchan in B. Tech Dept of CSE from NIT Durgapur, 3264 images constitute the comprehensive MRI dataset in total including 926 glioma tumors, 937 meningioma tumors, 901 pituitary tumors, and 500 no tumor.

Moreover, the performance of the models was assessed using five metrics: validation accuracy, test accuracy, precision, recall, and F1 score play pivotal roles in assessing the performance and generalization capabilities of a trained model.

Validation accuracy (Val acc) in Eq. 6 represents the model's precision on a separate dataset during training, measuring its ability to learn without overfitting the training data. By means of mathematics, validation accuracy is computed as:

$$Val\ acc = \frac{Total\ Number\ of\ instance\ in\ validation\ set}{Number\ of\ correctly\ predictec\ instrances} \tag{6}$$

Test accuracy (Test acc) in Eq. 7 reflects the model's proficiency in making accurate predictions on previously unseen data, providing insights into its real-world applicability. This metric is calculated by:

$$Test\ acc = \frac{Total\ number\ of\ instances\ in\ test\ set}{Number\ of\ correctly\ predicted\ instances} \tag{7}$$

Recall in Eq. 8, a metric crucial in scenarios where identifying true positives is paramount, is defined as:

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives} \tag{8}$$

Precision in Eq. 9 is a fundamental metric in the evaluation of classification models. Mathematically, precision is defined as the ratio of true positives (instances correctly predicted as positive) to the sum of true positives and false positives (instances incorrectly predicted as positive). The precision formula is given by:

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \tag{9}$$

The F1 score in Eq. 10, a harmonic mean of precision and recall, is expressed as:

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{10}$$

B. Scenario 1: The Results of Classifying MRI Images Into Two Classes: Normal or Glioma Tumor

The purpose of the experiments was to evaluate the effectiveness of the pre-trained models, after customization and training, in identifying the correct disease diagnosis of the MRI image. Moreover, these data can help us make easier and more intuitive comparisons across brain tumor categories (four classes or two classes).

Tables I and II show the performance evaluation metrics for classifying MRI images into normal or scoliosis. In transfer learning, the Resnet-50 achieved the highest accuracy (i.e., 100%). After fine-tuning, three models including MobileNet, InceptionV3, and EfficientNetB3 produced results that exceeded expectations (i.e., 100% for three models). In contrast, Resnet-50 showed worse results than the previous experiment.

TABLE I. The Accuracy of Classifying MRI Images Into Two Classes: Glioma Tumor and Normal in Transfer Learning, for Each Deep Learning Model

Model	**Transfer learning**				
	Val acc	Test acc	Precision	Recall	F1
EfficientNetB3	99.30%	98.60%	98.66%	98.60%	98.61%
ResNet50	100.00%	98.60%	98.60%	98.60%	98.60%
VGG16	99.30%	97.90%	98.02%	97.90%	97.92%
Ours	**99.30%**	**97.90%**	**97.92%**	**97.90%**	**97.91%**
InceptionV3	99.30%	93.01%	93.11%	93.01%	93.04%

TABLE II. The Accuracy of Classifying MRI Images Into Two Classes: Glioma Tumor and Normal in Fine Tuning, for Each Deep Learning Model

Model	**Fine tuning**				
	Val acc	Test acc	Precision	Recall	F1
EfficientNet B3	100.00%	99.30%	99.31%	99.30 %	99.30%
ResNet50	99.30%	96.50%	96.64%	96.50 %	96.53%
VGG16	99.30%	87.41%	88.33%	87.41 %	86.83%
Ours	**100.00%**	**97.90%**	**97.92%**	**97.90 %**	**97.91%**
InceptionV3	100.00%	98.60%	98.60%	98.60 %	98.60%

Fig. 5 and Fig. 6 show a sample training and validation progress curve showing the loss and accuracy values of ours in fine-tuning. The figure displays a stable learning behavior and appropriate training and validation sets.

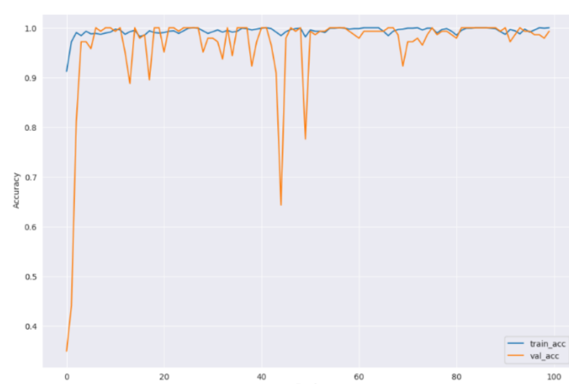

Fig. 5. Training accuracy and validation accuracy in fine-tuning of ours model (Glioma tumors and normal).

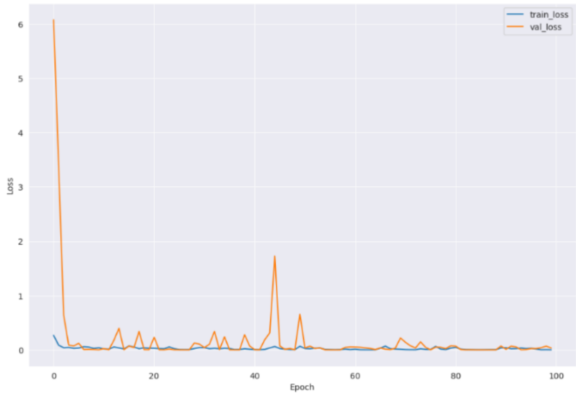

Fig. 6. Training loss in and validation loss in fine-tuning of ours model (Glioma tumors and normal).

Fig. 7 shows the confusion matrix for a sample run of ours model in two classes. In that run, the number of testing images is 143.

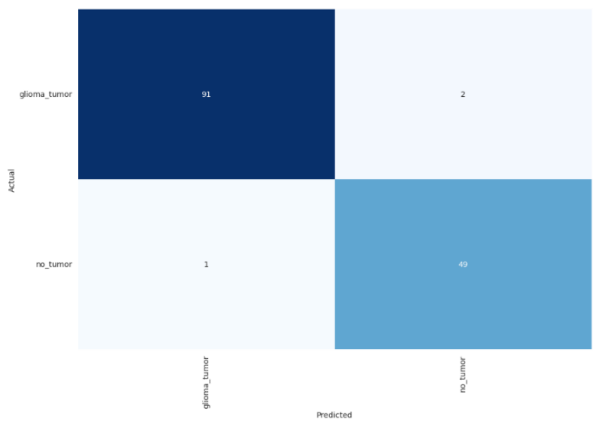

Fig. 7. Confusion matrix in fine tuning for ours model (Glioma tumors and normal).

C. *Scenario 2: The Results of Classifying MRI Images Into Two Classes: Normal or Meningioma Tumor*

Tables III and IV showed similar results in scenario 2 with the Resnet-50 achieving the biggest accuracy (i.e., 100%) in transfer learning. After fine-tuning, ours achieved extremely impressive results with low test loss and all other aspects reaching 100%.

TABLE III. THE ACCURACY OF CLASSIFYING MRI IMAGES INTO TWO CLASSES: MENINGIOMA TUMOR AND NORMAL IN TRANSFER LEARNING, FOR EACH DEEP LEARNING MODEL

Model	**Transfer learning**				
	Val acc	Test acc	Precision	Recall	F1
EfficientNetB3	99.31%	98.61%	98.64%	98.61%	98.60%
ResNet50	100.00%	96.53%	96.84%	96.53%	96.56%
VGG16	97.92%	97.92%	97.92%	97.92%	97.91%
Ours	**97.22%**	**95.14%**	**95.17%**	**95.14%**	**95.15%**
InceptionV3	95.83%	97.22%	97.43%	97.22%	97.25%

TABLE IV. THE ACCURACY OF CLASSIFYING MRI IMAGES INTO TWO CLASSES: MENINGIOMA TUMOR AND NORMAL IN FINE TUNING, FOR EACH DEEP LEARNING MODEL

Model	**Fine tuning**				
	Val acc	Test acc	Precision	Recall	F1
EfficientNetB 3	100.00%	99.31%	99.32%	99.31%	99.31%
ResNet50	100.00%	86.11%	90.08%	86.11%	86.44%
VGG16	99.31%	95.14%	95.17%	95.14%	95.15%
Ours	**100.00 %**	**100.00 %**	**100.00 %**	**100.00 %**	**100.00 %**
InceptionV3	100.00%	97.22%	97.22%	97.22%	97.22%

In Fig. 8 and Fig. 9, we can see the detail of the training accuracy and training loss about two classes that is pituitary tumors and normal MRI images. This chart can show us the high result of a classification of brain tumors when using ours model.

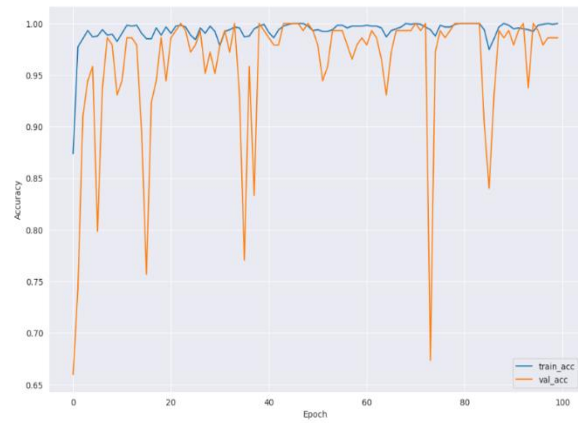

Fig. 8. Training accuracy and validation accuracy in fine-tuning of ours model (Meningioma tumors and normal).

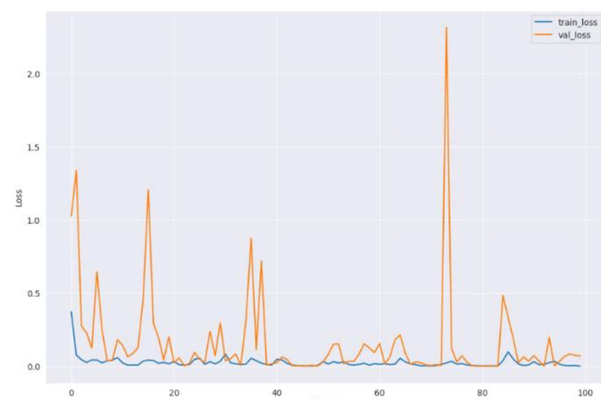

Fig. 9. Training loss in and validation loss in fine-tuning of ours model (Meningioma tumors and normal).

The confusion matrix in Fig. 10 shows that the result of model has a high performance when using it to classify brain tumor.

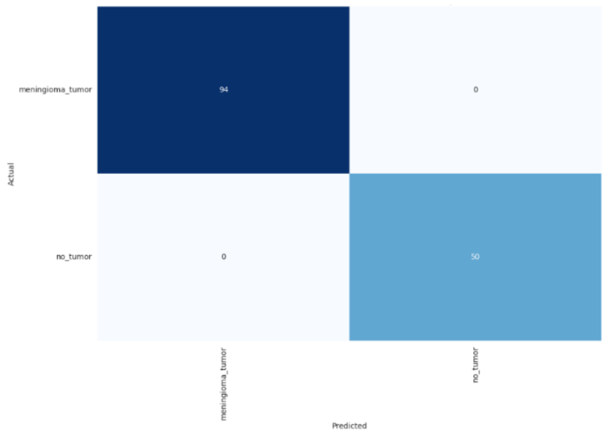

Fig. 10. Confusion matrix in fine tuning for ours model (Meningioma tumors and normal).

D. Scenario 3: The Results of Classifying MRI Images Into Two Classes: Normal or Pituitary Tumor

In scenario 3, Tables V and VI show that ResNet-50 has top results in all aspects (i.e., 100%) while it decreased a litter bit after fine-tuning, other models have a bit of a transformation (i.e., ~99%) when finished two steps.

TABLE V. THE ACCURACY OF CLASSIFYING MRI IMAGES INTO TWO CLASSES: PITUITARY TUMOR AND NORMAL IN TRANSFER LEARNING, FOR EACH DEEP LEARNING MODEL

Model	Transfer learning				
	Val acc	Test acc	Precision	Recall	F1
EfficientNetB3	100.00%	98.58%	98.58%	98.58%	98.57%
ResNet50	100.00%	100.00%	100.00%	100.00%	100.00%
VGG16	94.29%	98.58%	98.58%	98.58%	98.58%
Ours	**100.00%**	**99.29%**	**99.30%**	**99.29%**	**99.29%**
InceptionV3	99.29%	98.58%	98.58%	98.58%	98.58%

TABLE VI. THE ACCURACY OF CLASSIFYING MRI IMAGES INTO TWO CLASSES: PITUITARY TUMOR AND NORMAL IN FINE TUNING, FOR EACH DEEP LEARNING MODEL

Model	Fine tuning				
	Val acc	Test acc	Precision	Recall	F1
EfficientNetB3	100.00%	99.29%	99.30%	99.29%	99.29%
ResNet50	100.00%	99.29%	99.30%	99.29%	99.29%
VGG16	94.29%	35.46%	12.57%	35.46%	18.57%
Ours	**100.00%**	**99.29%**	**99.30%**	**99.29%**	**99.29%**
InceptionV3	100.00%	99.29%	99.30%	99.29%	99.29%

In this experiment, Fig. 11 and Fig. 12 give an explanation of training accuracy and training loss in two classes of normal and pituitary tumors in the 'Ours' model.

Finally, the result confusion matrix is represented in Fig. 13 which shows that the performance of the 'Ours' model is very successful.

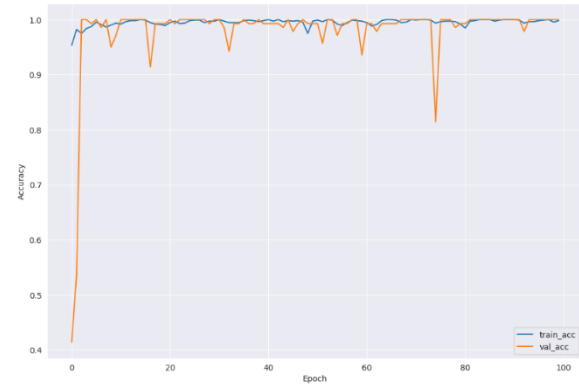

Fig. 11. Training accuracy and validation accuracy in fine-tuning of ours model (Pituitary tumors and normal).

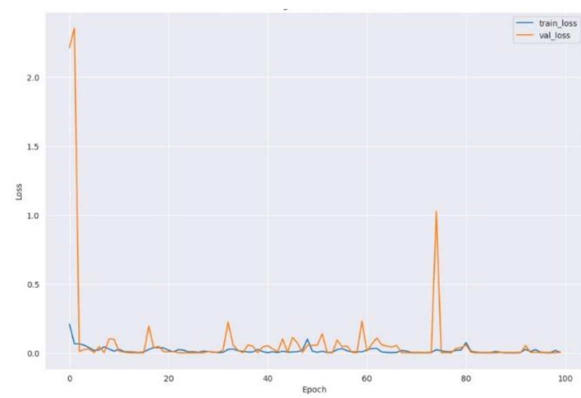

Fig. 12. Training loss in and validation loss in fine-tuning of ours model (Pituitary tumors and normal).

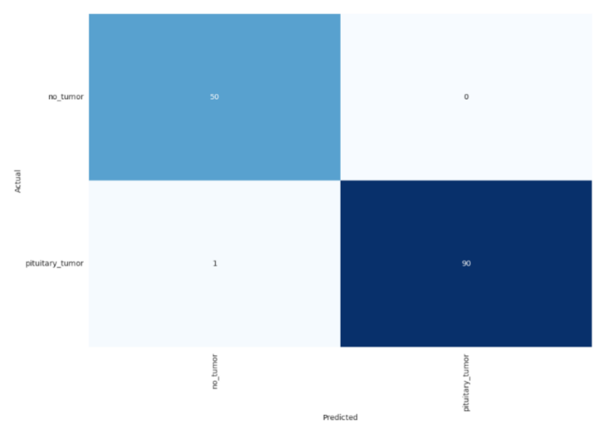

Fig. 13. Confusion matrix in fine tuning for ours model (Pituitary tumors and normal).

E. Scenario 4: The Results of Classifying MRI Images into Four Classes: Normal, Glioma Tumor, Meningioma Tumor, and Pituitary Tumor

Tables VII and VIII show the performance evaluation metrics for classifying MRI images into normal, glioma tumor, meningioma tumor, and pituitary tumor. The ResNet50 achieved the highest accuracy value in transfer learning over the three statistical measures with a validation accuracy of 94,17%, test accuracy of 92.05%, and F1 score of 92.06%. On the other hand, ours model performed the best after fine-tuning with a validation accuracy of 97,24%, test accuracy of 97.86%, and F1 score of 97.86%. The other performance metrics display a consistent and homogenous ability to identify negative as well as positive cases with a similar performance pattern to the accuracy results (i.e., Ours model achieving the best results). The significance of the F1 score lies in its role as an evaluation metric specifically designed for classification problems. An F1 score serves as an indicator of the model's accuracy, emphasizing its ability to achieve both high precision and recall.

TABLE VII. THE ACCURACY OF CLASSIFYING MRI IMAGES INTO FOUR CLASSES: NORMAL, GLIOMA TUMOR, MENINGIOMA TUMOR, AND PITUITARY TUMOR IN TRANSFER LEARNING, FOR EACH DEEP LEARNING MODEL

Model	Transfer learning				
	Val acc	Test acc	Precision	Recall	F1
EfficientNetB3	93.55%	91.13%	91.50%	91.13%	91.16%
ResNet50	94.17%	92.05%	92.18%	92.05%	92.06%
VGG16	86.50%	86.85%	87.09%	86.85%	86.88%
Ours	**89.88%**	**85.93%**	**86.27%**	**85.93%**	**85.89%**
InceptionV3	83.13%	81.65%	81.79%	81.65%	81.69%

TABLE VIII. THE ACCURACY OF CLASSIFYING MRI IMAGES INTO FOUR CLASSES: NORMAL, GLIOMA TUMOR, MENINGIOMA TUMOR, AND PITUITARY TUMOR IN FINE TUNING, FOR EACH DEEP LEARNING MODEL

Model	Fine tuning				
	Val acc	Test acc	Precision	Recall	F1
EfficientNetB3	97.55%	97.55%	97.62%	97.55%	97.55%
ResNet50	97.24%	95.11%	95.20%	95.11%	95.09%
VGG16	65.03%	15.29%	2.34%	15.29%	4.06%
Ours	**97.24%**	**97.86%**	**97.91%**	**97.86%**	**97.86%**
InceptionV3	97.24%	97.55%	97.56%	97.55%	97.55%

In Fig. 14 and Fig. 15 shows the training and validation progress curve for a sample run of the highest-performing model, which gives an indication of the fitting performance of the model and the need for more training. Training accuracy measures a model's performance on the training data, reflecting its ability to learn from the provided examples. Validation accuracy assesses the model's generalization to new, unseen data, helping identify potential overfitting or underfitting issues. Training loss quantifies the disparity between predicted and actual values in the training set, guiding the model to minimize errors during training. Validation loss mirrors this process on a separate dataset, serving as a key indicator of the model's generalization performance.

Fig. 16 shows the 'Ours' sample confusion matrix for four-class classification. This important step makes it possible for us to see a more intuitive comparison of the results achieved. Fig. 17 shows a sample output from the four-class classification process with the visual explanation by GradCam.

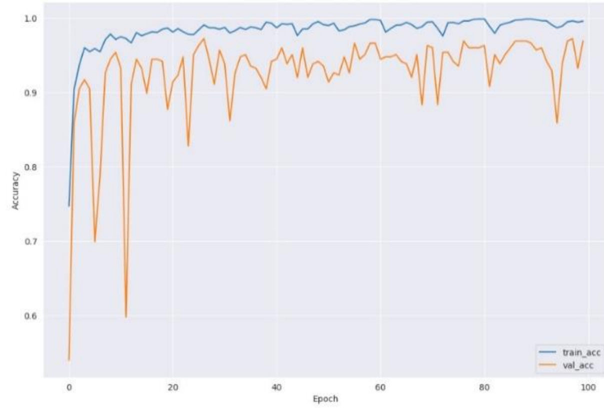

Fig. 14. Training accuracy and validation accuracy in fine-tuning of 'Ours' model (Four classes).

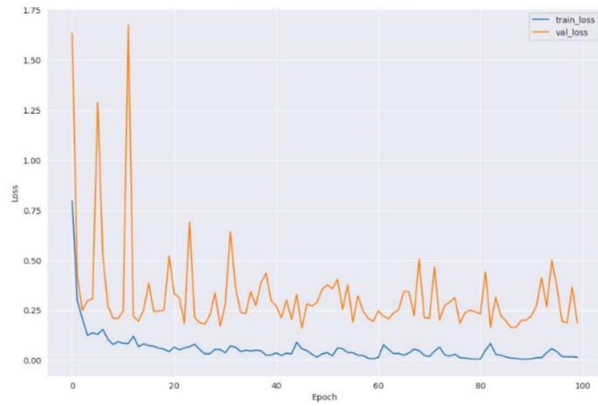

Fig. 15. Training loss in and validation loss in fine-tuning of 'Ours' model (Four classes).

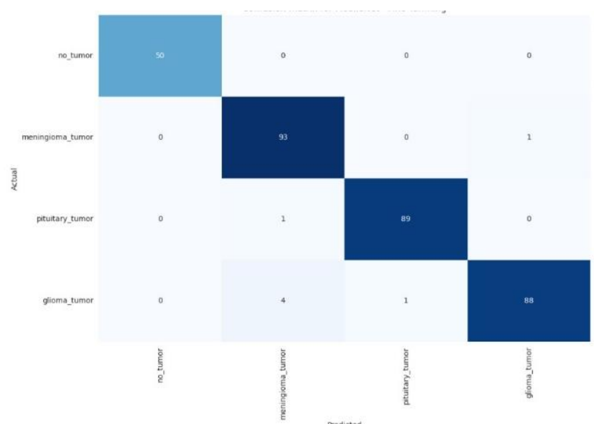

Fig. 16. Confusion matrix in fine tuning for 'Ours' model (Four classes).

F. Comparison with others State-of-the-art Methods

To examine the accuracy of the proposed model that our article has just given out in the previous section, we compare the accuracy score of the proposed model with other CNN architectures, which are VGG16, ResNet-50, ResNet-101and DenseNet201.

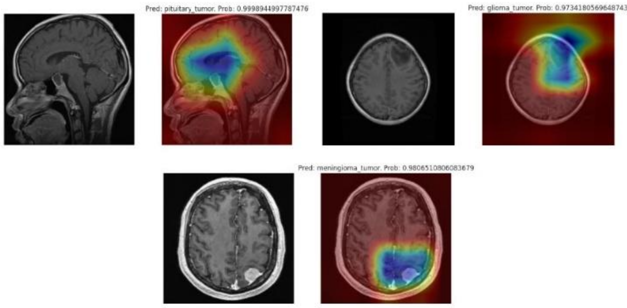

Fig. 17. Output of four classes classification.

Accuracy, precision, recall, and the F1 score offer many aspects of performance. Accuracy gives overall correctness but may mislead in imbalanced datasets. Precision focuses on the accuracy of positive predictions. Recall assesses the model's ability to capture all relevant positives. The F1 score strikes a balance between precision and recall, making it invaluable for tasks with uneven class distributions. Comparing these metrics involves weighing trade-offs based on task-specific priorities. While accuracy provides a broad overview, precision and recall cater to nuanced aspects, and the F1 score harmonizes their interaction, ensuring a well-informed evaluation of classification models within the constraints of particular objectives. Finally, the result of getting the value of training and accuracy on the test set is illustrated as in Table IX.

TABLE IX. COMPARISON WITH OTHERS STATE-OF-THE-ART METHODS

Ref.	***Proposed***	***Accuracy***
Wadhah Ayadi, et al.	CNN (ReLu)	94.74%
Ahmad Saleh, et al.	MobileNet	97.25%
Wadhah Ayadi, et al.	SVM	90.27%
Muhammad Imran Sharif,et al.	Densenet201	>95%
Sharan Kumar, et al.	Dolphin-SCA	96.30%
Díaz-Pernas, et al.	Multiscale CNN	97.30%
Huong Hoang Luong, et al.	ResNet-50	96%
Rinesh Sahadevan, et al.	MFNN	96.46%
Khan Swati	VGG19	94.82%
Nyoman Abiwinanda	Multiscale CNN	84.19%
Proposed model		**97.86%**

V. CONCLUSION

In our project, we utilized transfer learning, a powerful technique in machine learning, to enhance our model's performance in identifying brain tumors from MRI images. Transfer learning involves leveraging knowledge gained from a pre-trained model on a large dataset for a specific task and applying it to a different, but related, task. In our case, we used the MobileNet model, which was pre-trained on a vast dataset, as a starting point. This allowed our model to inherit knowledge about general image features, enabling it to focus on the intricacies of brain tumor classification.

Fine-tuning played a crucial role in tailoring the pre-trained MobileNet model to our specific medical imaging task. We incorporated dense and dropout layers while adjusting various hyperparameters to optimize the model's performance. The addition of these layers facilitated better feature extraction and prevented overfitting, contributing to the remarkable validation accuracy of 97.24%, test accuracy of 97.86%, and an F1 score of 97.86%.

To provide transparency and insights into our model's decision-making process, we adopted GradCam for visual explanations. This not only aids medical professionals in understanding the model's predictions but also accelerates medical examinations and treatments, making them more efficient and cost-effective.

On the other hand, while our project has shown promising results, certain drawbacks warrant consideration. One significant limitation is the size of the dataset used for training the model. The availability of a relatively small dataset can hinder the model's ability to generalize effectively to diverse and unseen cases. To address the issue of a small dataset, a potential solution involves acquiring and incorporating a more extensive and diverse set of MRI images for training. Collaborating with multiple medical institutions to aggregate data or exploring the use of data augmentation techniques could help augment the dataset, providing the model with a richer understanding of the variations in brain tumor presentations.

Additionally, the current model may face challenges in precisely identifying the boundaries of tumors, potentially leading to false positives or negatives. Furthermore, the reliance on a pre-trained MobileNet model, while beneficial for leveraging general image features, may introduce biases or limitations in capturing subtle nuances unique to medical images. Developing a custom architecture tailored to the intricacies of medical imaging, perhaps through architecture search techniques, could lead to a more specialized and optimized model for brain tumor classification.

In the future, the future trajectory involves refining data preparation, adopting advanced visualization methods, and expanding the dataset. We plan to enhance our model by refining our data preparation techniques, employing advanced visualization methods, and expanding our dataset. By doing so, we aim to further increase the accuracy and robustness of our model, reinforcing its role as a valuable tool in the medical field for the accurate and prompt classification of brain tumors in MRI scans. Our ongoing efforts underscore the significance of artificial intelligence in advancing medical diagnostics and treatment processes.

VI. ACKNOWLEDGMENT

This acknowledgment would be incomplete without expressing our sincere appreciation for Huong Hoang Luong, Hao Van Tran, and Phuc Tan Huynh's encouragement, support, suggestions, and positive influence. I am fortunate to have had the opportunity to collaborate with them, and I am deeply thankful for their contributions. Thank you, for your invaluable guidance and support. Your generosity and commitment have been crucial in bringing this research to fruition.

REFERENCES

[1] Q. T. Ostrom, M. Price, C. Neff, G. Cioffi, K. A. Waite, C. Kruchko, and J. S. Barnholtz-Sloan, "CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2016—2020," Neuro-Oncology, Oxford University Press US, vol. 25, no. Supplement_4, pp. iv1–iv99, 2023.

[2] J. Huang, H. Li, H. Yan, F.-X. Li, M. Tang, and D.-L. Lu, "The comparative burden of brain and central nervous system cancers from 1990 to 2019 between China and the United States and predicting the future burden," Frontiers in Public Health, Frontiers Media SA, vol. 10, p. 1018836, 2022.

[3] Q. T. Ostrom, N. Patil, G. Cioffi, K. Waite, C. Kruchko, and J. S. Barnholtz-Sloan, "CBTRUS statistical report: primary brain and other central nervous system tumors diagnosed in the United States in 2013–2017," Neuro-Oncology, Oxford University Press US, vol. 22, no. Supplement_1, pp. iv1–iv96, 2020.

[4] Q. T. Ostrom, M. Price, C. Neff, G. Cioffi, K. A. Waite, C. Kruchko, and J. S. Barnholtz-Sloan, "CBTRUS statistical report: primary brain and other central nervous system tumors diagnosed in the United States in 2015–2019," Neuro-Oncology, Oxford University Press US, vol. 24, no. Supplement_5, pp. v1–v95, 2022.

[5] G. Varoquaux and V. Cheplygina, "Machine learning for medical imaging: methodological failures and recommendations for the future," NPJ Digital Medicine, Nature Publishing Group UK London, vol. 5, no. 1, p. 48, 2022.

[6] M. Injadat, A. Moubayed, A. B. Nassif, and A. Shami, "Machine learning towards intelligent systems: applications, challenges, and opportunities," IEEE Transactions on Artificial Intelligence Review, Springer, vol. 54, pp. 3299–3348, 2021.

[7] X. Yu, J. Wang, Q.-Q. Hong, R. Teku, S.-H. Wang, and Y.-D. Zhang, "Transfer learning for medical image analyses: A survey," IEEE Transactions on Neural Networks and Learning Systems, Elsevier, vol. 489, pp. 230–254, 2022.

[8] E. Radiya-Dixit and X. Wang, "How fine can fine-tuning be? Learning efficient language models," in Proceedings of the International Conference on Artificial Intelligence and Statistics, PMLR, pp. 2435–2443, 2020.

[9] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A survey of convolutional neural networks: analysis, applications, and prospects," IEEE Trans. Neural Netw. Learn. Syst., 2021.

[10] W. Ayadi, W. Elhamzi, and M. Atri, "A new deep CNN for brain tumor classification," in 2020 20th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering (STA), IEEE, pp. 266–270, 2020.

[11] A. Saleh, R. Sukaik, and S. S. Abu-Naser, "Brain tumor classification using deep learning," in 2020 International Conference on Assistive and Rehabilitation Technologies (iCareTech), IEEE, pp. 131–136, 2020.

[12] W. Ayadi, I. Charfi, W. Elhamzi, and M. Atri, "Brain tumor classification based on hybrid approach," The Visual Computer, vol. 38, no. 1, pp. 107-117, 2022.

[13] M. I. Sharif, M. A. Khan, M. Alhussein, K. Aurangzeb, and M. Raza, "A decision support system for multimodal brain tumor classification using deep learning," Complex & Intelligent Systems, pp. 1-14, 2021.

[14] S. Kumar and D. P. Mankame, "Optimization driven deep convolution neural network for brain tumor classification," Biocybernetics and Biomedical Engineering, vol. 40, no. 3, pp. 1190-1204, 2020.

[15] F. J. Díaz-Pernas, M. Martínez-Zarzuela, M. Antón-Rodríguez, and D. González-Ortega, "A deep learning approach for brain tumor classification and segmentation using a multiscale convolutional neural network," in Healthcare, vol. 9, no. 2, p. 153, 2021.

[16] H. T. Nguyen, H. H. Luong, P. T. Phan, H. H. D. Nguyen, D. Ly, D. M. Phan, and T. T. Do, "HS-UNET-ID: An approach for human skin classification integrating between UNET and improved dense convolutional network," International Journal of Imaging Systems and Technology, vol. 32, no. 6, pp. 1832-1845, 2022.

[17] H. H. Luong, N. H. Khang, N. Q. Le, D. M. Canh, P. S. Ha, et al., "A Proposed Approach for Monkeypox Classification," International Journal of Advanced Computer Science and Applications, vol. 14, no. 8, 2023

[18] H. T. Nguyen, H. H. Luong, T. H. N. Kien, N. P. L. Phan, T. D. Thuan, T. T. Tin, T. D. Nguyen, and T. C. Toai, "Brain Tumors Detection on MRI Images with K-means Clustering and Residual Networks," in International Conference on Computational Collective Intelligence, pp. 317-329, 2022.

[19] S. Rinesh, K. Maheswari, B. Arthi, P. Sherubha, A. Vijay, S. Sridhar, T. Rajendran, Y. A. Waji, et al., "Investigations on brain tumor classification using hybrid machine learning algorithms," Journal of Healthcare Engineering, vol. 2022, 2022.

[20] Z. N. K. Swati, Q. Zhao, M. Kabir, F. Ali, Z. Ali, S. Ahmed, and J. Lu, "Brain tumor classification for MR images using transfer learning and fine-tuning," Computerized Medical Imaging and Graphics, vol. 75, pp. 34-46, 2019.

[21] N. Abiwinanda, M. Hanif, S. T. Hesaputra, A. Handayani, and T. R. Mengko, "Brain tumor classification using convolutional neural network," in World Congress on Medical Physics and Biomedical Engineering 2018: June 3-8, 2018, Prague, Czech Republic (Vol. 1), pp. 183-189, 2019.

[22] H. H. Luong, L. T. T. Le, H. T. Nguyen, V. Q. Hua, K. V. Nguyen, T. N. P. Bach, T. N. A. Nguyen, and H. T. Q. Nguyen, "Transfer learning with fine-tuning on MobileNet and Grad-CAM for bones abnormalities diagnosis," in Computational Intelligence in Security for Information Systems Conference, pp. 171-179, 2022.

[23] H. H. Luong, N. T. L. Phan, T. C. Dinh, T. M. Dang, T. T. Duong, T. D. Nguyen, and H. T. Nguyen, "Fine-Tuning MobileNet for Breast Cancer Diagnosis," in Inventive Computation and Information Technologies: Proceedings of ICICIT 2022, pp. 841-856, 2023.

[24] H. H. Luong, P. T. Vo, H. C. Phan, N. L. D. Tran, H. Q. Le, and H. T. Nguyen, "Fine-Tuning VGG16 for Alzheimer's Disease Diagnosis," in Conference on Complex, Intelligent, and Software Intensive Systems, pp. 68-79, 2023.